

# Rajkumar Pujari

email: [rajkumarsaikorian@gmail.com](mailto:rajkumarsaikorian@gmail.com)    [rajkumar-pujari.com](http://rajkumar-pujari.com)    Github: [@pujari-rajkumar](https://github.com/pujari-rajkumar)    Phone: +1 (765)-701-7617

EDUCATION	<b>Purdue University</b> , West Lafayette, Indiana, USA Ph.D. in Natural Language Processing, Department of Computer Science	May 2025 3.8/4
	<b>Indian Institute of Technology Kharagpur</b> , West Bengal, India B.Tech. (Hons.), Department of Computer Science and Engineering	May 2014 9.1/10
RELEVANT WORK EXPERIENCE	<b>Applied Scientist II</b> Amazon AGI, Sunnyvale, CA	Aug 2025–Present
	Post-training Nova models. Experience with building and owning LLM-as-a-judge for data filtering, data re-writing, RL, and evaluation. Worked on building preference and rubric following LLM judges. Worked on hill-climbing model’s general response style and quality. Extensive experience with distributed training at scale (SFT, RL, DPO). Also worked on building synthetic data generation and cleaning pipelines.	
RELEVANT INTERNSHIPS	<b>JEM Intern, Microsoft Research, Redmond</b> <i>RL-guided Multitask model for Stereotype Detection in Text</i>	May 2021–Aug 2021 E Nouri, E Oveson and P Kulkarni
	Worked on Reinforcement Learning guided Multitask Learning Model for <i>Stereotype Detection</i> . Leveraged several toxicity datasets such as <i>Hate-Speech</i> , <i>Misogyny &amp; Offensive Language</i> to systematically improve performance on the task of <i>Stereotype Detection</i> . Built a <i>Stereotype Detection</i> dataset using Amazon Mechanical Turk.	
	<b>Applied Scientist Intern, Amazon Alexa Conversational Search Team</b> <i>Conversational Question Answering in Sports Domain</i>	May 2019–Aug 2019 Kevin Small
	Developed a new conversational QA dataset for sports domain. Built a ParlAI-style data collection tool and collected the dataset using Amazon MTurk. Also designed a generative neural architecture for the task.	
PUBLICATIONS	<b>LLM-Human Schema Pipeline for Cultural Context Grounding of Conversations</b> Rajkumar Pujari and Dan Goldwasser Talk at NAACL 2025 main conference, Albuquerque, USA April 29–May 4	
	<b>“We Demand Justice!”: Towards Social Context Grounding of Political Texts</b> Rajkumar Pujari, Chengfei Wu and Dan Goldwasser Talk at EMNLP 2024 main conference, Miami, USA November 12–16	
	<b>Reinforcement Learning Guided Multi-Task Learning for Low-Resource Stereotype Detection</b> Rajkumar Pujari, Erik Oveson, Priyanka Kulkarni and Elnaz Nouri Poster at ACL 2022 main conference, Dublin, Ireland from May 22–27	
	<b>Understanding Politics via Contextualized Discourse Processing</b> Rajkumar Pujari and Dan Goldwasser Talk at EMNLP 2021 main conference, Punta Cana, Dominican Republic on November 6–12	
	<b>Using Natural Language Relations between Answer Choices for Machine Comprehension</b> Rajkumar Pujari and Dan Goldwasser Talk at NAACL-HLT 2019 main conference, Minneapolis, USA on June 2–7	
	<b>Can Taxonomy help? Improving Semantic Question Matching using Question Taxonomy</b> Deepak Gupta, Rajkumar Pujari, Asif Ekbal, Pushpak Bhattacharyya, Anutosh Maitra, Tom Jain and Shubhashis Sengupta Poster at COLING 2018 main conference, Santa Fe, New Mexico, USA, August 20–24	
	<b>A Novel Two-stage Framework for Extracting Opinionated Sentences from News Articles</b> Rajkumar Pujari, Swara Desai, Niloy Ganguly and Pawan Goyal Talk at Texgraphs-9 workshop at EMNLP 2014, Doha, Qatar, October 25–29	
RESEARCH EXPERIENCE	<b>Ph.D. Student, Purdue University</b> <i>Contextualized Discourse Representations; Machine Comprehension; Social Norm Understanding (Pragmatics)</i> Prof. Dan Goldwasser	Aug 2017–May 2025
	Worked on investigating the usefulness of common sense knowledge and inferences in machine comprehension. Worked on developing a distributed, interpretable representation for large political corpora and also, working on a DARPA CCU (Computational Cultural Understanding) project that leverages Large Language Models combined with symbolic models to extract social norms and their violations from conversations.	

	<b>Project Research Assistant, CFILT, IIT Bombay</b> <i>Semantic Question matching using a taxonomy and DL representations</i> Developed a restricted-domain QA system for Accenture labs, Bangalore. Was primarily focused on developing a hierarchical taxonomy and subsequent classification algorithms used for semantic question matching.	Jan 2016–May 2017 Prof. Pushpak Bhattacharyya
POSITIONS OF RESPONSIBILITY	<b>Program Committee Member - Reviewer; *Outstanding Reviewer</b> ACL ARR since Nov 2021, ACL (2022–2024), ACL-IJCNLP 2021 *, EMNLP (2020–2024), EMNLP-IJCNLP 2019, NAACL-HLT (2019–2021), NAACL (2022–2024), AACL (2020–2023), COLM 2024, COLING (2025, 2022, 2020), LREC-COLING 2024, IJCAI (2024, 2023, 2020), TALLIP  <b>Graduate Teaching Assistant</b> Data Mining and Machine Learning, Web Information Search and management, Data structures and algorithms, and Operating Systems undergraduate courses.  <b>President, Purdue University Cricket Club</b> Organized cricket tournaments for 250 players, raised funds, and managed an annual budget of USD \$7000.	Jul 2017–Present  May 2019–Jun 2021
OTHER WORK EXPERIENCE	<b>Senior Quantitative Researcher</b> WorldQuant Research, India Was responsible for researching financial and mathematical literature, understanding various datasets to identify sources of market inefficiencies and convert them to predictive profitable models called <i>alphas</i> . The objective was to identify and construct signals and make robust models from them with high Sharpe ratios (returns/risk) and significant abnormal returns. Concentrated mainly on seeking low turnover quality <i>alphas</i> for trading in the equity market which are used in developing algorithmic daily re-balancing long-short trading strategies on US, Europe, Asian, and other markets	Jul 2014–Dec 2015
NOTABLE ACHIEVEMENTS	Was promoted to Senior Quantitative Researcher after 1 <sup>st</sup> year at Worldquant Research for exceptional performance Secured a Department Change from Electrical Engineering department to Computer Science department at the end of the first year based on academic merit (5% acceptance rate)  Secured 840 <sup>th</sup> rank amongst more than 450,000 students (99.81 percentile) in IIT-JEE 2010 and 939 <sup>th</sup> rank amongst more than 1,000,000 students (99.91 percentile) in AIEEE 2010  Qualified among Center Top 10% in National Standard Examination in Physics and National Standard Examination in Chemistry conducted by Indian Association of Physics Teachers (IAPT) in Class XII standard.	
KEY COURSES	<b>Graduate Level:</b> Advanced Topics In Reasoning with LLMs (A+), ML Methods for NLP, Statistical ML (A+), Deep Learning, Numerical Methods for Optimization, Algorithm Design and Analysis, Operating Systems <b>Data Mining:</b> Information Retrieval, Machine Learning, Speech and Natural Language Processing <b>Electives:</b> Advanced Graph Theory, Artificial Intelligence, Computational Number Theory, Database Management and Systems, Distributed Systems, Formal Systems, Foundations of Cryptography	
KEY PROJECTS	<b>Summer Project, IIT Kharagpur</b> <i>Tracking soccer players in a video</i> Worked on various color segmentation algorithms and implemented a tracking algorithm based on position prediction in successive frames using MATLAB Image Processing Toolkit	May 2012–Jun 2012 Prof.Partha Pratim Das
EXTRA-CURRICULAR ACTIVITIES	Represented Purdue Cricket Club in MWCC (2018) and MCT (2019, 2021). Senior Writing Team Member of Entrepreneurship Cell IIT Kharagpur in the academic year 2011–2012 National Service Scheme (NSS) volunteer in the Education Improvement Group and a National Cadet Corps (NCC) cadet for 6 years and was awarded NCC-A & B certificates	
REFERENCES	Prof. Dan Goldwasser, Purdue University Leora Morgenstern, SRI Kevin Small, Amazon.com	email: <a href="mailto:dgoldwas@purdue.edu">dgoldwas@purdue.edu</a> email: <a href="mailto:leora.morgenstern@sri.com">leora.morgenstern@sri.com</a> email: <a href="mailto:smakevin@amazon.com">smakevin@amazon.com</a>